

The Multivariate Revolution in Marketing Research

JAGDISH N. SHETH

Can the current multivariate methods revolution in marketing research be explained? What is the role of computer technology in the rapid diffusion of multivariate methods? This article defines multivariate analysis and discusses the reasons for the probable continuing increase in its use in marketing research.

Journal of Marketing, Vol. 35 (January, 1971), pp. 13-19.

MANY would agree with the statement that the computer has produced significant advances in the natural and social sciences. However, this general observation overlooks the fact that these two areas have applied computer technology in different ways.

The current diffusion of computer technology is occurring at a time when most of the natural sciences possess several well-developed and invariant laws based on deductive reasoning. Under these circumstances the computer has provided opportunities for model building and for programming a complex network of constructs which enables large-scale testing of physical laws. The most outstanding example of these applications has been provided by the successful exploration of outer space.

However, the social sciences, including marketing, have yet to develop invariant laws. The result is that most of the research in this area is empirical. Attempts are made to explore realities in order to understand the basic nature of the disciplines. Thus, since much of marketing research is empirical and exploratory, the computer has been primarily used to analyze, sort, process, and compact standard commercial data into manageable data banks.

Perhaps computer utilization for model building in the natural sciences and for data analysis in the social sciences provides the best indication of the anticipated rapid adoption of multivariate methods in marketing research. In addition, two facilitating conditions have emerged which ensure large-scale diffusion of multivariate methods in the future.

The first condition refers to the fact that after three decades of systematic data-gathering, marketing research has learned the art of data collection. For example, procedures exist for drawing accurate samples from populations, training interviewers and respondents, receiving cooperation from respondents, designing structured questionnaires, and coding and tabulating collected data. In this respect, the marketing discipline may be more advanced than some of the other social sciences such as political science. In fact, the increasing accumulation and storage of market research reflects the validity and usefulness of the information collected. Today it is difficult to find a large-scale enterprise which has not been affected by the information explosion.

Second, the market place is a complex phenomenon. A multitude of factors intervene between the marketing activities of companies and market responses. A simple input-output approach does not seem to provide satisfactory answers to marketing problems. Therefore, attempts are constantly made to examine intervening factors and how they mediate between marketing activities and market responses. This has resulted in the collection of information which corresponds to the complexity of the phenomenon.

The capability of the computer to process these complex, large-scale data banks has resulted in the increased use of multivariate methods in marketing research. The extent of this "multivariate revolution" in marketing research is indicated by several factors. For example, a vast number of canned computer programs for these techniques are already developed and available.¹ In addition, several reviews on the usages of multivariate methods in marketing have been written.² A third indication is provided by the increasing number of articles in such journals as the *JOURNAL OF MARKETING*, *Journal of Marketing Research*, and *Journal of Advertising Research* which treat applications of multivariate methods to marketing problems.

Inevitably, some questions may be raised about this revolution: How long will it last? Is it not just another fad which will fade away as soon as a new one is introduced? What will be the consequences on future marketing research if multivariate methods are here to stay? Which techniques will be the most relevant and important? However, before these questions can be answered, existing multivariate methods should be understood and classified.

What is Multivariate Analysis?

Although Kendall gives a more technical definition,³ it is possible to characterize multivariate analysis as all statistical methods which simultaneously analyze more than two variables on a sample of observations. As such these methods are extensions of univariate analysis (all known distributions including binomial, poisson, and normal distribution as well as probability system and Bayesian approaches to the analysis of one variable), and bivariate analysis (including cross-classification, correlation, and simple regression used to analyze two variables).

Figure 1 presents a classification of most of the multivariate methods. It is based on three judgments the marketing researcher must make about the nature and utilization of his data: (1) Are some of the variables dependent upon others, thereby requiring special treatment? (2) If yes, how many are to be treated as dependent in a single

analysis? and (3) What are the presumed properties of the data? Specifically, are the data *qualitative* (nonmetric) in that the marketing reality is scaled on nominal or ordinal scales, or *quantitative* (metric) and scaled on interval or ratio scales? The technique to be utilized will depend upon the answers to these three questions.

Multiple regression, including several of its variants such as stepwise regression and simultaneous regression, is the appropriate method of analysis when the researcher has a single, metric dependent variable which is presumed to be a function of other independent variables. The objective of multiple regression is to predict the variability in the dependent variable based on its covariance with all the independent variables. This objective is then achieved by the statistical rule of least squares.

Whenever the researcher is interested in predicting the level of the dependent phenomenon, he would find multiple regression useful. For example, sales are predicted from the knowledge of their past relationship (covariance) with marketing efforts; market shares have been predicted based on consumer preference, retail structure, and point-of-purchase advertising and promotion; and consumer buying behavior is predicted from the knowledge of personality and socioeconomic profiles.

If the single dependent variable is dichotomous (e.g., male-female) or multichotomous (e.g., high-medium-low), and therefore nonmetric, the multivariate method of *multiple discriminant analysis* is appropriate. The primary objective of discriminant analysis is to *predict* an entity's likelihood of belonging to a particular class or group based on several predictor variables. For example, discriminant analysis has been widely used in marketing to predict whether (1) a person is a good or poor credit risk based on his socioeconomic and demographic profile, (2) innovators can be distinguished from noninnovators according to their psychological and socioeconomic profiles, and (3) private label buyers can be separated from national brand buyers based on socioeconomic and purchasing differences.⁴

The primary objective of multiple discriminant analysis is to correctly classify entities into mutu-

⁴ Sheth, same reference as footnote 2.

¹ Kenneth M. Warwick, "Computerized Multivariate Methods," paper presented at AMA Workshop on Multivariate Methods in Marketing, Chicago, January 21-23, 1970.

² Jagdish N. Sheth, "Multivariate Analysis in Marketing," *Journal of Advertising Research*, Vol. 10 (February, 1970), pp. 29-39; Ronald E. Frank and Paul E. Green, "Numerical Taxonomy in Marketing Analysis: A Review Article," *Journal of Marketing Research*, Vol. 5 (February, 1968), pp. 83-98; and Paul E. Green, Frank J. Carmone, and Patrick J. Robinson, "Nonmetric Scaling Methods: An Exposition and Overview," *Wharton Quarterly*, Vol. 2 (Winter-Spring, 1968), pp. 159-173.

³ Maurice G. Kendall, *A Course in Multivariate Analysis* (London: Charles Griffin & Company, 1957).

• ABOUT THE AUTHOR. Jagdish N. Sheth is presently acting head and associate professor of business administration at the University of Illinois. He was formerly on the faculty of Columbia University and Sloan School of Management, M.I.T. In 1969, Professor Sheth was program chairman of the workshop in Multivariate Methods in Marketing sponsored by the American Marketing Association. His research and publications are primarily in consumer psychology, multivariate methods, and international marketing.



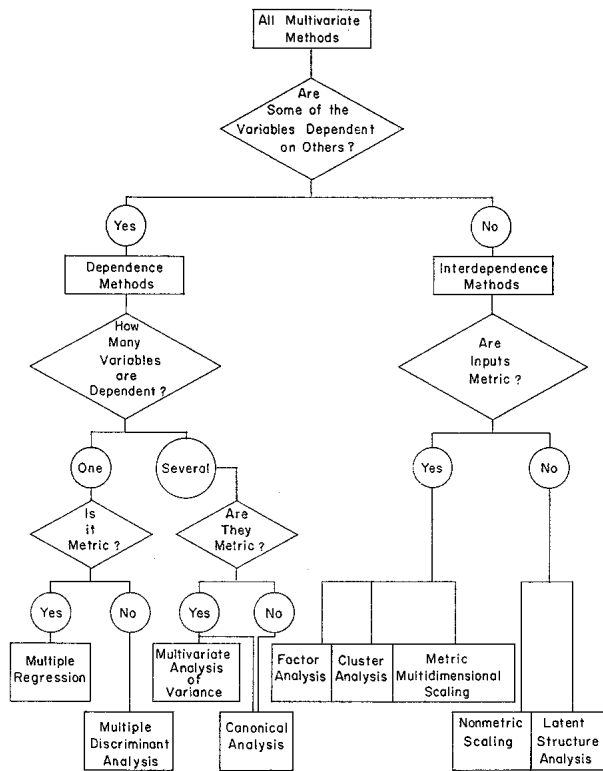


FIGURE 1. A classification of multivariate methods.

ally exclusive and exhaustive classes or groups. This objective is achieved by the statistical decision rule of maximizing the ratio of among-group to within-group variance-covariances on the profile of discriminating (predictor) variables. In addition to the prediction of class membership based on the profile, discriminant analysis reveals which specific variables in the profile account for the largest proportion of intergroup differences.

Multivariate analysis of variance (multi-ANOVA) is an extension of bivariate analysis of variance in which the ratio of among-groups variance to within-groups variance is calculated on a set of variables instead of a single variable. As such, multi-ANOVA is useful whenever the researcher is testing hypotheses concerning multivariate differences in group responses to experimental manipulations. For example, he may be interested in using one test market and one control market to examine the effect of

an advertising campaign on sales as well as awareness, knowledge, and attitudes.

The objective in *canonical analysis* is to simultaneously predict a set of dependent variables from their joint covariance with a set of independent variables. Both metric and nonmetric data are acceptable in canonical analysis. The procedure followed is to obtain a set of weights for the dependent and independent variables which provides the maximum simple correlation between the composite dependent variable and the composite independent variable.

Canonical analysis appears very useful in marketing because the multitude of marketing and environmental factors tend to produce a variety of market responses. The writer, for example, is currently investigating the "hierarchy of effects" (awareness-interest-attitude-conviction-action) as multiple consequences of advertising and promotion.

Thus far the discussion has focused on multivariate methods applied to data which contain both dependent and independent variables. However, if the researcher is investigating the interrelations, and therefore the interdependence, among all the variables, several other multivariate methods are appropriate. These include factor analysis, cluster analysis, and metric multidimensional scaling if the variables are presumed to be metric, and nonmetric multidimensional scaling and latent structure analysis if they are presumed to be nonmetric.

Factor analysis is based on this proposition: If there is systematic interdependence among a set of observed (manifest) variables, it must be due to something more fundamental (latent) which creates this commonality. One can even consider all the manifest variables as simply *indicators* of this fundamental factor. What is this factor? Can it be extracted from the observed data and their relationships? Is it unidimensional or multidimensional? For example, can an individual's income, education, occupation, and dwelling area be considered as indicators of his social class? How can this factor be extracted? Conversely, factor analysis is also used as a data-reduction method which summarizes the commonality of all the manifest variables into a few factors.

The statistical approach utilized in factor analysis is to maximally summarize all of the variance (information), including the covariance (interdependence), in as few factors as possible, while retaining the flexibility of reproducing the original relationship among the manifest variables.

Factor analysis has been widely used in marketing. It has been used to (1) extract latent dimensions of relative liquor preferences such as sweetness, price, and regional popularity; (2) cluster a series of night-time television programs or magazines based on their relative viewership and readership; and (3) systematically search for powerful predictors of a phenomenon under investigation.⁵

In *cluster analysis*, the objective is to classify a population of entities into a small number of mutually exclusive and exhaustive groups based on the similarities of profiles among entities. Unlike discriminant analysis, the groups are not predefined. In fact, two major objectives are to determine *how many* groups really exist in the population, and what is their composition.

Cluster analysis seems useful for market segmentation on personality, socioeconomic and demographic, psychological, and purchasing characteristics of the consumers. However, several other applications have been made in marketing. Examples include the clustering of test market cities in order that they may be selected and controlled for experimentation, and grouping a variety of computers based on their objective performance characteris-

tics.⁶ Most of the clustering methods are judgmental, however, and devoid of statistical inferences. In fact, judgment is needed in selecting and coding attributes, in obtaining indices of resemblance or similarity, in choosing among various clustering algorithms, and in naming and testing derived clusters.

Both *metric* and *nonmetric multidimensional scaling* methods, unlike all other multivariate methods, start with a single piece of information. This information relates to perceived relative similarities among a set of objects, such as products, from a sample of respondents. The basic assumption in both metric and nonmetric multidimensional scaling methods is that people perceive a set of objects as being more or less similar to one another on a number of dimensions (usually uncorrelated with one another) instead of only one. However, it may be impossible to directly obtain this multidimensional map from the respondents. One reason for this difficulty is that the respondent may not be consciously aware that he is judging similarities among objects based on these dimensions. A second reason is that he is unwilling to reveal factors (dimensions) that enter into his judgment on similarities. Given this impossibility of directly obtaining the dimensions, reliance is placed on statistical methods of multidimensional scaling to infer the number and types of dimensions that presumably underlie the expressed relative similarities among objects. Therefore, multidimensional scaling methods are applicable in those areas of marketing research where *motivation research* is currently used.

In both metric and nonmetric multidimensional scaling, the judged similarities among a set of objects (e.g., products, suppliers) are statistically transformed into distances by placing those objects in a multidimensional space of some dimensionality. For example, if objects A and B are perceived by the respondent as being most similar compared to all other possible pairs of objects, these techniques will position objects A and B in such a way that the distance between them in multidimensional space is shorter than that between any two other objects.

Despite the similarities between metric and nonmetric multidimensional scaling, there are two important differences. First, metric multidimensional scaling extracts the dimensionality of metric similarity data, whereas the input to nonmetric multidimensional scaling is nonmetric (ordinal) similarities. The metric similarities are often obtained on a bipolar similarity scale on which pairs of objects are rated one at a time. The nonmetric similarities are obtained by asking respondents to rank order (on the basis of similarity) all possible pairs that can be obtained from a set of objects. Various procedures such as dyadic or triadic combinations or rating scales can be used. Second, metric multidimensional scaling attempts to reduce the observed

⁵ Sheth, same reference as footnote 2.

⁶ Frank and Green, same reference as footnote 2.

similarities to be represented in a space of minimum dimensions, from the trivial representation in $n-1$ dimensions, where n is the number of objects. In nonmetric multidimensional scaling, the objective is to metricize the nonmetric data by transforming nonmetric data into a metric space, and then by reducing the dimensionality. This is done by a decision rule of monotone transformation in which the observed rank orderings of pairs of objects are reproduced as closely as possible in an arbitrary metric space of some specified dimensions. This metric space is usually Euclidian, although non-Euclidian spaces can be created by the computer.

Although metric multidimensional scaling has not been applied to any large extent in marketing, nonmetric multidimensional scaling has become very popular in the last three years under the pioneering efforts of Paul Green. It has been applied to the dimensionality of similarities among automobiles, magazines, graduate schools, and several other sets of objects.⁷

Latent structure analysis shares both of the objectives of factor analysis: to extract latent factors and express relationship of manifest variables with these factors as their indicators, and to classify a population of respondents into pure types. Traditionally, nonmetric data have been the input to latent structure analysis, although recently metric data have also been used. In marketing, the only applications of this method have been by Myers and Nicosia.⁸ One of the main reasons for this has been the lack of computer programs to handle the tedious calculations inherent in this method.

Is Multivariate Revolution a Fad?

A number of compelling reasons suggest that the rapid use of multivariate methods in marketing is not a fad. Instead, these methods are so important that they will be used more frequently in the future.

First, let us examine the anatomy of several behavioral and operations research methods (e.g., pupil dilation, Markov chains) that degenerated into fads. This was due to three factors. First, some operations research methods clearly proved to be ahead of their times. They presumed (through model building) considerable knowledge about the response functions to marketing efforts at a time when no one actually understood how the marketing mix is related to market reactions. These research methods may prove useful once some laws of market behavior have been established. Second, other behavioral and operations research methods took a normative posture of how markets *may* or should

behave at a time when the empirical inductive approach of descriptive research was considered more appropriate. Third, some methods, particularly in the behavioral area, proved to be genuine fads because they created the illusion that market complexity can be easily described by simple "buzz word" models.

None of these factors seems to be present in multivariate methods. Multivariate methods are largely empirical, deal with the market reality by working backward from reality to conceptualization, and easily handle the complexity presumed to be inherent in marketing research.

Second, multivariate methods as "innovative methods" seem to be consistent with modern marketing concepts of focusing on marketing research needs. And the most pressing need of marketing research is the ability to analyze complex data. This is clearly indicated by the following statement: "For the purposes of marketing research or any other applied field, most of our tools are, or should be, multivariate. One is pushed to a conclusion that unless a marketing problem is treated as a multivariate problem, it is treated superficially."⁹

As discussed earlier, this need for complex analysis is manifested today since data collection is a well-developed and standardized art, and computer capabilities are easily accessible.

Finally, a number of multivariate methods are simply extensions of univariate and bivariate analysis of data. Also, there exist a great variety of multivariate methods. Both of these factors contribute toward inhibiting their degeneration to fadism, because fads generally involve highly specialized research tools. However, some specific multivariate techniques may become fads due to overselling. In addition, we should expect the usual problems of coordinating the man-machine interface which are inevitable in the use of multivariate methods.

However, none of these factors is likely to deter the progress of the multivariate revolution primarily because all the facilitating conditions are present today.

Role of the Computer in Multivariate Revolution

Perhaps the most important factor in the rapid diffusion of multivariate methods in marketing research is the availability of computer programs. In fact, we can assert that the lack of computer programs has been a major factor in the imbalance between the extensive data banks in existence today and their weak statistical analysis in most marketing research activities. It would seem that a union between multivariate methods and the computer will provide excellent opportunities for more scientific approaches to marketing problems.

⁹ Ronald Gatty, "Multivariate Analysis for Marketing Research: An Evaluation," *Applied Statistics*, Vol. 15 (November, 1966), p. 158.

⁷ Green, Carmone, and Robinson, same reference as footnote 2.

⁸ John G. Myers and Francesco M. Nicosia, "New Empirical Directions in Market Segmentation: Latent Structure Models," in *Changing Marketing Systems: Consumer, Corporate and Government Interfaces*, Reed Moyer (ed.) (Chicago: American Marketing Association, 1967).

What are the effects of this union on the development of marketing information systems? At present, marketing information systems in most companies basically consist of large data banks. However, a truly useful marketing information system requires an integrated approach between data banks and their retrieval and analysis in accordance with the needs of marketing management. Since most management decisions are complex, a truly multivariate analysis is needed that can be undertaken only if computer facilities are readily available. For example, a recurring decision in marketing management will involve budget allocation among several marketing forces, including advertising, direct mail, promotion, and personal selling. Ramond and Sheth have developed a marketing information system for budget allocation in which time-series audit data on market responses and marketing activities are analyzed by multivariate regression.¹⁰ In essence, changes in shares of market responses such as sales are regressed on changes in shares of several marketing forces including advertising, direct mail, and promotion. Their relative weights and signs are then used by the marketing manager to choose one of the following alternatives given that his objective is to increase the profitability of marketing forces: (1) Maintain the present budget allocation policy. (2) Increase the total budget by a certain amount to reach the optimum level of profitability. (3) Reallocate the budget among marketing forces proportionate to their relative weights. (4) Reduce the total budget by a certain amount to bring expenditures to the optimum level. (5) Phase out the product. This type of marketing information system could not be achieved without a complete interface between the computer and some multivariate method.

A second area benefiting from this interface is testing and estimating parameters of complex and comprehensive theories of the market place. Two specific examples may illuminate this point. First, in the area of advertising effectiveness, a number of researchers¹¹ have conceptualized a "hierarchy of effects" of advertising and promotion. This hierarchy usually begins with awareness and ends with purchase behavior; in between, several other effects such as interest, knowledge, preference, liking, and conviction are sequentially arranged. It is also presumed that advertising will have, in general, less impact as we move from awareness to action. It would seem that despite numerous empirical studies, no study has as yet attempted to validate

the hierarchy by using an appropriate multivariate method.¹² Since the theory presumes a number of effects, canonical analysis appears most appropriate to test the theory and estimate parameters of relative relationships between the hierarchy of effects and a set of advertising variables such as media and copy. Unless such a complex multivariate analysis is done, it is not possible to either support or reject the theory of multiple advertising effectiveness. Much of the inconclusive support currently found in the research literature is perhaps due to this lack of multivariate analysis. Such multivariate analysis, however, was impossible without the appropriate computer capabilities.

Another example comes from an outstanding effort by Farley and Ring to fully test the Howard-Sheth theory of buyer behavior through the use of simultaneous linear equations and the computer.¹³ Howard and Sheth have developed a comprehensive and complex theory of buyer behavior in which a large number of psychological constructs, such as attention, overt search, attitude, motives, and satisfaction, intervene between the marketing stimuli and the buyer's responses. In addition, a number of exogenous factors, such as social class, culture, and reference groups, also determine a buyer's responses via their influence on the psychological constructs. Finally, several of the constructs are dynamically interdependent on one another because of the theory's information processing framework. Farley and Ring operationally defined these interdependencies in terms of a set of eleven simultaneous equations; then, using the panel data collected as part of the Columbia Buyer Behavior Project on a sample of more than 900 respondents, they tested the theory. Although they were only moderately successful in validating the theory, their effort represents one of the best examples of how the union between the computer and multivariate methods facilitates the testing of complex theories.

There are several areas of marketing research in which only univariate data have been collected, although the phenomenon is recognized to be complex. In these areas, the combination of multivariate methods and the computer may be most beneficial in furthering systematic and scientific analysis to possibly generate some invariant laws. An example is the research on durable appliances, particularly related to purchasing plans of households. Despite the recognition that purchasing plans are determined by a composite of several important factors, most attempts at measuring them

¹⁰ Charles Ramond and Jagdish Sheth, "Controlling Marketing Performance: Two Case Examples," paper presented to the Workshop on Marketing Information Systems, Association of National Advertisers, August 20, 1970.

¹¹ Robert C. Lavidge and Gary A. Steiner, "A Model for Predictive Measurement of Advertising Effectiveness," *JOURNAL OF MARKETING*, Vol. 25 (October, 1961), pp. 59-62.

¹² Kristian S. Palda, "The Hypothesis of a Hierarchy of Effects: A Partial Evaluation," *Journal of Marketing Research*, Vol. 3 (February, 1966), pp. 13-24.

¹³ John V. Farley and L. Winston Ring, "Deriving an Empirically Testable Version of the Howard-Sheth Theory of Buyer Behavior," paper presented at the Third Annual Buyer Behavior Conference, Columbia University, May, 1969.

have remained univariate. A single scale is used on which the degree of certainty of buying intentions within a specified time period is obtained from the respondents. It is very probable that this univariate scale is used as a surrogate for more complex factors and has not represented the construct well enough to either predict or explain subsequent purchasing behavior. With the use of multivariate methods such as factor analysis, it is conceivable that buying intentions may indeed prove to be a multidimensional concept.

Conclusion

A number of facilitating factors suggest that multivariate methods may rapidly diffuse in marketing research, and may become a way of life in the statistical analysis of marketing data. These include (1) the empirical, inductive tendency in conducting marketing research due to lack of discovery

of marketing laws; (2) collection of large-scale data on marketing problems; (3) confidence in data banks in terms of their reliability and validity; and (4) availability of computers and canned computer programs. The last factor is certainly the most important one in enhancing the diffusion of multivariate methods.

The role of the computer in furthering the maturity of the marketing discipline is thus immense. By diffusing multivariate methods, it is likely to enable marketing researchers to attempt large-scale marketing information systems in which an integrated marketing approach can be undertaken. It will enable researchers to test and estimate parameters of complex generalized theories and models. With the use of multivariate methods, the computer is likely to generate a sudden increase in in-depth scientific empirical research on well-known issues in marketing.